

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

JC979 U.S. PTO  
10/036718  
12/21/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日  
Date of Application:

2000年12月25日

出 願 番 号  
Application Number:

特願2000-391855

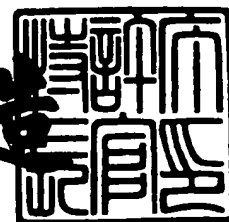
出 願 人  
Applicant(s):

松下電器産業株式会社

2001年 9月19日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

及 川 耕 造



CERTIFIED COPY OF  
PRIORITY DOCUMENT

出証番号 出証特2001-3086320

【書類名】 特許願

【整理番号】 2900825610

【提出日】 平成12年12月25日

【あて先】 特許庁長官殿

【国際特許分類】 H03M 7/30

【発明者】

【住所又は居所】 神奈川県横浜市港北区綱島東四丁目3番1号 松下通信  
工業株式会社内

【氏名】 渡邊 泰仁

【特許出願人】

【識別番号】 000005821

【氏名又は名称】 松下電器産業株式会社

【代理人】

【識別番号】 100072604

【弁理士】

【氏名又は名称】 有我 軍一郎

【電話番号】 03-3370-2470

【手数料の表示】

【予納台帳番号】 006529

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9908698

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声符号化方式、音声符号化方法およびそれを用いる音声符号化装置、記録媒体、ならびに音楽配信システム

【特許請求の範囲】

【請求項 1】 入力信号を複数のスケールファクタバンドに分割するとともに、所定の聴覚心理モデルに基づいて前記スケールファクタバンドごとに符号化を行う音声符号化方式であって、

前記入力信号が定常的な信号であるか過渡的な信号であるかにより前記入力信号の符号化フレームを特定する窓長を分析する窓長分析手段と、

前記入力信号を高速フーリエ変換する F F T 分析手段と、

符号化モードを決定するための符号化モード情報を指定する符号化モード情報指定手段と、

前記入力信号を高速フーリエ変換した結果および前記聴覚心理モデルに基づいて、前記入力信号に対する信号対マスク比を算出する聴覚心理分析手段と、

符号化する周波数帯域を特定する最大スケールファクタバンドを算出する最大スケールファクタバンド算出手段と、

前記最大スケールファクタバンドを算出するための係数を格納した最大スケールファクタバンド係数テーブルを記憶する最大スケールファクタバンドテーブル記憶手段と、

前記窓長および前記符号化モード情報に基づいて、前記最大スケールファクタバンドを算出するために必要な初期値を算出する最大スケールファクタバンド初期値算出手段と、

前記 F F T 分析手段からのデータおよび前記最大スケールファクタバンドの算出値に基づいて前記入力信号のスペクトル処理を行うスペクトル処理手段と、

前記スペクトル処理がされたデータを前記スケールファクタバンドごとに量子化および符号化する量子化／符号化手段とを備え、

前記入力信号に応じて適応的に前記最大スケールファクタバンドを算出するようにしたことを特徴とする音声符号化方式。

【請求項 2】 前記最大スケールファクタバンドは、前記符号化モード情報

から得られるビットレートおよびサンプリング周波数に応じ、前記最大スケールファクタバンド係数テーブルを参照することにより設定されることを特徴とする請求項 1 に記載の音声符号化方式。

【請求項 3】 前記最大スケールファクタバンドは、前記符号化モード情報より得られるチャンネル数に応じ、前記最大スケールファクタバンド係数テーブルを参照することにより設定されることを特徴とする請求項 2 に記載の音声符号化方式。

【請求項 4】 前記窓長および前記符号化モード情報に応じた所定の複数の閾値を格納した閾値テーブルを記憶する閾値テーブル記憶手段を有し、

前記最大スケールファクタバンド算出手段は、

前記最大スケールファクタバンド係数テーブルを参照することにより設定されるいずれか 1 つの最大スケールファクタバンドにおける、前記聴覚心理分析手段で算出された信号対マスク比と、前記窓長および前記符号化モード情報に応じて前記閾値テーブルを参照することにより設定された閾値とを比較し、

前記算出された信号対マスク比が前記設定された閾値より小さい場合には前記最大スケールファクタバンドを 1 ずつ下げていき、前記算出された信号対マスク比が前記設定された閾値より大きな値をとるスケールファクタバンドになる 1 つ前のスケールファクタバンドを現在の符号化フレームの最大スケールファクタバンドとすることを特徴とする請求項 1 に記載の音声符号化方式。

【請求項 5】 前記窓長および前記符号化モード情報に応じた所定の複数の閾値を格納した閾値テーブルを記憶する閾値テーブル記憶手段を有し、

前記最大スケールファクタバンド算出手段は、

前記最大スケールファクタバンド係数テーブルを参照することにより設定されるいずれか 1 つの最大スケールファクタバンドにおける、前記入力信号のスペクトルのエネルギー値を算出し、その算出値と前記窓長および前記符号化モード情報に応じて前記閾値テーブルを参照することにより設定された閾値とを比較し、前記エネルギー値が前記設定された閾値より小さい場合には前記最大スケールファクタバンドを下げていき、前記エネルギー値が前記設定された閾値より大きな値をとるスケールファクタバンドになる 1 つ前のスケールファクタバンドを現符

号化フレームの最大スケールファクタバンドとすることを特徴とする請求項 1 に記載の音声符号化方式。

【請求項 6】 前記所定の閾値は、前記符号化モード情報のうちビットレートを参照することにより決定されることを特徴とする請求項 4 または 5 に記載の音声符号化方式。

【請求項 7】 前記所定の閾値は、前記符号化モード情報のうち入力のスプリング周波数をも参照することにより決定されることを特徴とする請求項 6 に記載の音声符号化方式。

【請求項 8】 前記所定の閾値は、前記符号化フレームの窓長をも参照することにより決定されることを特徴とする請求項 7 に記載の音声符号化方式。

【請求項 9】 前記所定の閾値は、前記符号化モード情報のうちチャンネル数をも参照することにより決定されることを特徴とする請求項 7 または 8 に記載の音声符号化方式。

【請求項 10】 前記窓長分析手段により前記入力信号が過渡的であることが分析され、ビットレートが所定値より高く、前記スケールファクタバンドのそれぞれに割付け可能なビット数が所定値を超えて十分にある状態において、前記最大スケールファクタバンド算出手段が算出した最大スケールファクタバンドが所定の最小スケールファクタバンドより低い場合には、最大スケールファクタバンドの設定値を算出した値より所定の増分だけ上げることを特徴とする請求項 1 に記載の音声符号化方式。

【請求項 11】 サンプリングした入力信号のデータを複数の周波数帯域のスケールファクタバンドに分割して、所定の聴覚心理モデルに基づき、符号化する周波数帯域を制限しながら、その周波数帯域内のスケールファクタバンドごとに符号化を行う音声符号化方法であって、

前記入力信号が定常的な信号であるか過渡的な信号であるかにより前記入力信号の符号化フレームを特定する窓長を分析する窓長分析ステップと、

前記窓長および符号化モードの指定情報に基づいて前記スケールファクタバンドの周波数帯域順に複数のスケールファクタバンドを設定するとともに、所定の閾値の情報に基づいて、前記スケールファクタバンドのうち前記符号化する周波

数帯域を特定する最大スケールファクタバンドを初期設定する初期設定ステップと、

前記入力信号の周波数分布を分析した結果および前記聴覚心理モデルに基づいて、前記最大スケールファクタバンドを算出する最大スケールファクタバンド算出ステップと、

前記最大スケールファクタバンドにより制限された周波数帯域内で、前記スケールファクタバンドごとのスペクトル処理を実行するスペクトル処理ステップと

前記スペクトル処理後のデータを量子化および符号化処理する符号化処理ステップとを含むことを特徴とする音声符号化方法。

【請求項 1 2】 前記最大スケールファクタバンド算出ステップに先立って前記入力信号の周波数分布を分析するとともに、前記スケールファクタバンドごとに前記聴覚心理モデルに基づく信号対マスク比を算出し、その信号対マスク比と前記所定の閾値とに基づいて前記最大スケールファクタバンドを決定することを特徴とする請求項 1 1 に記載の音声符号化方法。

【請求項 1 3】 前記最大スケールファクタバンド算出ステップに先立って前記入力信号の周波数分布を分析するとともに、前記スケールファクタバンドごとにスペクトルのエネルギー値を算出し、そのエネルギー値と前記閾値とに基づいて前記最大スケールファクタバンドを決定することを特徴とする請求項 1 1 に記載の音声符号化方法。

【請求項 1 4】 請求項 1 1 乃至請求項 1 3 のいずれかに記載の音声符号化方法を用いて音声符号化を行うこと特徴とする音声符号化装置。

【請求項 1 5】 請求項 1 1 乃至請求項 1 3 のいずれかに記載の音声符号化方法の各ステップを実現するためのプログラムが記録されていることを特徴とするコンピュータ読取可能な記録媒体。

【請求項 1 6】 請求項 1 乃至請求項 1 5 に記載の音声符号化方式または請求項 1 4 に記載の音声符号化装置を構成要素とすることを特徴とする音楽配信システム。

【発明の詳細な説明】

## 【 0 0 0 1 】

## 【発明の属する技術分野】

本発明は、音声符号化方式、音声符号化方法およびそれを用いる音声符号化装置、記録媒体、ならびに音楽配信システムに関し、特にMPEG (Moving Picture Experts Group) オーディオ規格で採用されている時間周波数変換符号化を用いる音声符号化方式、音声符号化方法、音声符号化装置、および前記音声符号化方法を実行ためのプログラムを格納した記録媒体、ならびにその音声符号化方式または装置を構成要素とする音楽配信システムに関するものである。

## 【 0 0 0 2 】

## 【従来の技術】

従来、音声符号化には、統計的性質を利用して符号化を行うエントロピー符号化や、人間の知覚特性を利用して不要な細部情報を削減していく知覚符号化が利用されているが、MPEGオーディオ規格では後者が積極的に利用されている。例えば、人間の聴覚での最小可聴限界以下の音や、マスキング効果により聞き取れなくなる周波数帯域の音を省略する等した圧縮方法が採用されている。

## 【 0 0 0 3 】

この種のデジタルオーディオ信号を符号化する際に、符号化する周波数帯域を設定する場合は、入力信号のサンプリング周波数あるいは窓長（ロングフレーム・ショートフレーム）に応じて、固定の値を割り当てていた。ここで、MPEG 2オーディオ規格準拠の音声符号化方式を例にとり、その動作を説明する。図 1 8 は、MPEG 2オーディオ規格AAC (Advanced Audio Coding) 準拠の音声符号化方式における、最大スケールファクタバンド値 (max Sfb) の一例のデータを示すテーブルである。このテーブルは、入力信号のサンプリング周波数に対するロングフレーム、ショートフレームそれぞれについて最大スケールファクタバンド値を割り当てたものである。

## 【 0 0 0 4 】

図 1 9 は、従来例の音声符号装置のブロック図である。

この音声符号化装置は、FFT分析手段300、窓長分析手段310、入力信号情報320、聴覚心理分析手段330、最大スケールファクタバンド値（図中

では  $\max Sfb$  と記す) 算出手段 340、前記最大スケールファクタバンド値 ( $\max Sfb$ ) を格納したテーブル 350、スペクトル処理手段 360、および量子化／符号化手段 370 を含んで構成されている。

#### 【0005】

この音声符号化装置に入力されるデジタルオーディオ信号は、窓長分析手段 310 により分析され、入力信号が定常な信号の場合にはロングフレームと、入力信号が過渡的な信号の場合にはショートフレームと分析される。

この窓長分析手段 310 の分析結果は、符号化モードを指定する手段 320 より得られた入力信号のサンプリング周波数と共に、最大スケールファクタバンド算出手段 340 に取り込まれる。この最大スケールファクタバンド算出手段 340 は、入力内容に応じて最大スケールファクタバンド ( $\max Sfb$ ) 係数テーブル 350 を参照し、最大スケールファクタバンド値 ( $\max Sfb$ ) を設定する算出処理を実行する。

#### 【0006】

一方、入力デジタルオーディオ信号は、FFT 分析手段 300 において、高速フーリエ変換 (FFT: Fast Fourier Transform) され、その結果を用いて、聴覚心理分析手段 330 により人間の聴覚心理モデルに基づくスケールファクタバンドごとの信号対マスク比 (SMR; Signal to Masking Ratio) 値が算出される。そして、この信号対マスク比 (SMR) および最大スケールファクタバンド値 ( $\max Sfb$ ) 等の情報を用いて、スペクトル処理手段 360 が MDCT (Modified Discrete Cosine Transform; 変形離散コサイン変換) 処理や TNS (Temporal Noise Shaping; 時間領域雑音形状化処理) 等のスペクトル処理を実行し、次いで、処理量子化／符号化手段 370 が量子化／符号化を行うようになっている。

#### 【0007】

##### 【発明が解決しようとする課題】

しかしながら、以上のように構成された従来の音声符号化方式では、符号化する周波数帯域を特定する最大スケールファクタバンド ( $\max Sfb$ ) の値が予め設定された固定値であるため、例えば低周波数域にのみ信号が偏っているよう



な場合にも最大スケールファクタバンド (max Sfb) まで、演算処理および量子化／符号化処理を行うことになるため、演算量の増加および符号化に必要な符号量が増加することになっていた。そのため、特に音楽配信システムにおいて、無駄な符号化演算および符号量が多くなり、その結果として音質向上が容易でなかった。

#### 【0008】

本発明は、上記のような従来の問題点に鑑みてなされたもので、符号化する周波数帯域を入力信号の周波数特性に応じて適応的に変化させることにより、演算量を削減し、符号量を有効に活用することができる音声符号化方式、音声符号化方法およびその措置、その音声符号化方式を実現するためのプログラムを格納したコンピュータ読取可能な記録媒体、ならびに音楽配信システムを提供することを目的とする。

#### 【0009】

##### 【課題を解決するための手段】

上記課題を解決するため、本発明の音声符号化方式は、入力信号を複数のスケールファクタバンドに分割するとともに、所定の聴覚心理モデルに基づいて前記スケールファクタバンドごとに符号化を行う音声符号化方式であって、前記入力信号が定常的な信号であるか過渡的な信号であるかにより前記入力信号の符号化フレームを特定する窓長を分析する窓長分析手段と、前記入力信号を高速フーリエ変換するFFT分析手段と、符号化モードを決定するための符号化モード情報を指定する符号化モード情報指定手段と、前記入力信号を高速フーリエ変換した結果および前記聴覚心理モデルに基づいて、前記入力信号に対する信号対マスク比を算出する聴覚心理分析手段と、符号化する周波数帯域を特定する最大スケールファクタバンドを算出する最大スケールファクタバンド算出手段と、前記最大スケールファクタバンドを算出するための係数を格納した最大スケールファクタバンド係数テーブルを記憶する最大スケールファクタバンドテーブル記憶手段と、前記窓長および前記符号化モード情報に基づいて、前記最大スケールファクタバンドを算出するために必要な初期値を算出する最大スケールファクタバンド初期値算出手段と、前記FFT分析手段からのデータおよび前記最大スケールファ

クタバンドの算出値に基づいて前記入力信号のスペクトル処理を行うスペクトル処理手段と、前記スペクトル処理がされたデータを前記スケールファクタバンドごとに量子化および符号化する量子化／符号化手段とを備え、前記入力信号に応じて適応的に前記最大スケールファクタバンドを算出するようにしたことを特徴とするものである。この構成により、入力信号に応じて最大スケールファクタバンドを適応的に算出することで、符号化する周波数帯域を常に必要な帯域に割当てて、符号化のために無駄な演算量を削減し、符号量を有効に活用することができる音声符号化方式を容易に実現することができる。

上記音声符号化方式においては、前記最大スケールファクタバンドは、前記符号化モード情報から得られるビットレートおよびサンプリング周波数に応じ、前記最大スケールファクタバンド係数テーブルを参照することにより設定されるのが好ましい。この構成により、符号化モードに適した最大スケールファクタバンドの設定が可能となる。

#### 【0010】

また、前記最大スケールファクタバンドは、前記符号化モード情報より得られるチャンネル数に応じ、前記最大スケールファクタバンド係数テーブルを参照することにより設定されるのが、より好ましい。この構成により、入力信号のチャンネル数をも考慮して、その符号化モードにより適した最大スケールファクタバンドの設定が可能となる。

#### 【0011】

また、上記音声符号化方式においては、前記窓長および前記符号化モード情報に応じた所定の複数の閾値を格納した閾値テーブルを記憶する閾値テーブル記憶手段を有し、前記最大スケールファクタバンド算出手段は、前記最大スケールファクタバンド係数テーブルを参照することにより設定されるいずれか1つの最大スケールファクタバンドにおける、前記聴覚心理分析手段で算出された信号対マスク比と、前記窓長および前記符号化モード情報に応じて前記閾値テーブルを参照することにより設定された閾値とを比較し、前記算出された信号対マスク比が前記設定された閾値より小さい場合には前記最大スケールファクタバンドを1ずつ下げていき、前記算出された信号対マスク比が前記設定された閾値より大きな

値をとるスケールファクタバンドになる1つ前のスケールファクタバンドを現在の符号化フレームの最大スケールファクタバンドとするものであるのがよい。この構成により、マスキング閾および臨界帯域を常に好適に保って無駄な符号化処理を省き、符号化効率を高めることができる。

## 【 0 0 1 2 】

あるいは、上記音声符号化方式においては、前記窓長および前記符号化モード情報に応じた所定の複数の閾値を格納した閾値テーブルを記憶する閾値テーブル記憶手段を有し、前記最大スケールファクタバンド算出手段は、前記最大スケールファクタバンド係数テーブルを参照することにより設定されるいずれか1つの最大スケールファクタバンドにおける、前記入力信号のスペクトルのエネルギー値を算出し、その算出値と前記窓長および前記符号化モード情報に応じて前記閾値テーブルを参照することにより設定された閾値とを比較し、前記エネルギー値が前記設定された閾値より小さい場合には前記最大スケールファクタバンドを下げていき、前記エネルギー値が前記設定された閾値より大きな値をとるスケールファクタバンドになる1つ前のスケールファクタバンドを現符号化フレームの最大スケールファクタバンドとするものであってもよい。この構成によっても、マスキング閾および臨界帯域を常に好適に保って無駄な符号化処理を省き、符号化効率を高めることができる。

## 【 0 0 1 3 】

前記所定の閾値は、前記符号化モード情報のうちビットレートを参照することにより決定されるのがよい。さらに、前記所定の閾値は、前記符号化モード情報のうち入力のスプリング周波数をも参照することにより決定されるものであってもよく、前記符号化フレームの窓長をも参照することにより決定されるものであってもよい。また、前記所定の閾値は、前記符号化モード情報のうちチャンネル数をも参照することにより決定されるものであってもよい。このような構成により、符号化モードに応じてきめ細かに閾値を設定することができ、演算量の削減効果をより高めることができる。

## 【 0 0 1 4 】

また、上記音声符号化方式においては、前記窓長分析手段により前記入力信号

が過渡的であることが分析され、ビットレートが所定値より高く、前記スケールファクタバンドのそれぞれに割付け可能なビット数が所定値を超えて十分にある状態において、前記最大スケールファクタバンド算出手段が算出した最大スケールファクタバンドが所定の最小スケールファクタバンドより低い場合には、最大スケールファクタバンドの設定値を算出した値より所定の増分だけ上げるようにするのが好ましい。このような構成により、削減した周波数帯域の符号量（ビット割付量）を、音質に影響する周波数帯域の符号化に有効活用して、音質を向上させることができる。

## 【 0 0 1 5 】

一方、本発明の音声符号化方法は、サンプリングした入力信号のデータを複数の周波数帯域のスケールファクタバンドに分割して、所定の聴覚心理モデルに基づき、符号化する周波数帯域を制限しながら、その周波数帯域内のスケールファクタバンドごとに符号化を行う音声符号化方法であって、前記入力信号が定常的な信号であるか過渡的な信号であるかにより前記入力信号の符号化フレームを特定する窓長を分析する窓長分析ステップと、前記窓長および符号化モードの指定情報に基づいて前記スケールファクタバンドの周波数帯域順に複数のスケールファクタバンドを設定するとともに、所定の閾値の情報に基づいて、前記スケールファクタバンドのうち前記符号化する周波数帯域を特定する最大スケールファクタバンドを初期設定する初期設定ステップと、前記入力信号の周波数分布を分析した結果および前記聴覚心理モデルに基づいて、前記最大スケールファクタバンドを算出する最大スケールファクタバンド算出ステップと、前記最大スケールファクタバンドにより制限された周波数帯域幅内で、前記スケールファクタバンドごとのスペクトル処理を実行するスペクトル処理ステップと、前記スペクトル処理後のデータを量子化および符号化処理する符号化処理ステップとを含むことを特徴とする。この構成により、入力信号に応じて最大スケールファクタバンドを適応的に算出することで、符号化する周波数帯域を常に必要な帯域に割当てて、符号化のために無駄な演算量を削減し、符号量を有効に活用することができる。

## 【 0 0 1 6 】

上記音声符号化方法においては、前記最大スケールファクタバンド算出ステッ

プに先立って前記入力信号の周波数分布を分析するとともに、前記スケールファクタバンドごとに前記聴覚心理モデルに基づく信号対マスク比を算出し、その信号対マスク比と前記所定の閾値とに基づいて前記最大スケールファクタバンドを決定することが好ましい。この構成により、マスキング閾および臨界帯域を常に好適に保って無駄な符号化処理を省き、符号化効率を高めることができる。

## 【 0 0 1 7 】

また、前記最大スケールファクタバンド算出ステップに先立って前記入力信号の周波数分布を分析するとともに、前記スケールファクタバンドごとにスペクトルのエネルギー値を算出し、そのエネルギー値と前記閾値とに基づいて前記最大スケールファクタバンドを決定するのが好ましい。この構成により、マスキング閾および臨界帯域を常に好適に保って無駄な符号化処理を省き、符号化効率を高めることができる。

## 【 0 0 1 8 】

本発明の音声符号化装置は、上記音声符号化方法を用いて音声符号化を行うものであり、かかる構成により、入力信号に応じて最大スケールファクタバンドを適応的に算出することで、符号化する周波数帯域を常に好適にして無駄な演算量を削減し、符号量を有効に活用することができる音声符号化装置となる。

## 【 0 0 1 9 】

本発明の記録媒体は、上記音声符号化方法の各ステップを実現するためのプログラムが記録されているコンピュータ読取可能なものである。この構成により、記録媒体の読取が可能なハードウェア資源を用いて上記音声符号化方法の各ステップの処理を実行することができ、当該方法の効果を得ることができる。

## 【 0 0 2 0 】

本発明の音楽配信システムは、上記音声符号化方式または音声符号化装置を構成要素とするものであり、かかる構成によって、上記音声符号化方式または音声符号化装置による音声符号化を実行し、高符号化効率で音質の優れた音楽の配信を行うことができる音楽配信システムとなる。

## 【 0 0 2 1 】

## 【発明の実施の形態】

以下、本発明の好ましい実施の形態について、図面に基づき説明する。

(実施の形態 1)

本発明の実施の形態 1 にかかる音声符号化方式および装置について、図 1 ～図 8 を参照しながら説明する。

【0022】

本実施形態の音声符号化装置においては、図 1 に概略のブロック構成を示すように、デジタルオーディオ信号である入力信号は F F T 分析手段 1 0 0 に入力され、ここで高速フーリエ変換 (F F T : Fast Fourier Transform) され、入力信号の周波数分布が分析される。入力信号はまた窓長分析手段 1 1 0 にも入力され、ここでその入力信号が定常的な信号である場合にはロングフレームの入力信号ブロックと、その入力信号が過渡的な信号である場合にはショートフレームの入力信号ブロックと分析される。これにより、現在の入力信号に適した符号化フレームの長さが判定されることになる。

【0023】

また、入力信号のサンプリング周波数やビットレート等のような符号化に必要な情報は、符号化モードを指定する情報として符号化モード情報指定手段 1 2 0 により指定入力される。この指定入力は、予めユーザーが符号化モード情報指定手段 1 2 0 に操作入力することで行われる。

【0024】

F F T 分析手段 1 0 0 の出力は聴覚心理分析手段 1 3 0 に与えられ、この聴覚心理分析手段 1 3 0 により公知の人間の聴覚心理モデルに基づく、信号対マスク比 (S M R) の値を算出する処理がなされる。この S M R 値は、各周波数帯域のスケールファクタバンドごとに割り当てられるもので、各周波数領域のスケールファクタバンドの信号レベルとマスキングレベルの比の対数値として与えられるものである。

【0025】

一方、窓長分析手段 1 1 0 で分析された窓長と、符号化モード情報指定手段 1 2 0 に指定入力された情報、例えばサンプリング周波数は、最大スケールファクタバンド (m a x S f b) 初期値設定手段 1 4 0 に入力され、ここで最大スケ-

ルファクタバンド (max Sfb) の初期値が算出される。この最大スケールファクタバンド初期値設定手段 140 による算出処理は、最大スケールファクタバンド値 (max Sfb) 算出に必要な初期値および信号対マスク比 (SMR) 閾値を記録している最大スケールファクタバンド (max Sfb) 係数テーブル記憶手段 180 を参照しながら、後述する手順で実行される。なお、図 2 に示すように、最大スケールファクタバンド係数テーブル記憶手段 180 には、後述する最大スケールファクタバンド初期値テーブル 410 および信号対マスク比 (SMR) 閾値テーブル 420 が記憶されている。

#### 【0026】

最大スケールファクタバンド初期値設定手段 140 からの初期値情報と聴覚心理分析手段 130 により算出された SMR 値とは、それぞれ最大スケールファクタバンド適応算出手段 150 に入力され、ここで最大スケールファクタバンド (max Sfb) が入力信号の特性に応じて適応的に算出される。

#### 【0027】

そして、最大スケールファクタバンド適応算出手段 150 で算出された最大スケールファクタバンド (max Sfb) と聴覚心理分析手段 130 により算出された SMR 値に基づいて、スペクトル処理手段 160 で MDCT 処理や TNS 処理等のスペクトル処理が行われ、その処理後のデータが量子化／符号化手段 170 によって量子化され、符号化データとして出力される。

#### 【0028】

なお、上記のような音声符号化装置の構成は、例えばサウンドカードを内蔵したパーソナルコンピュータの ROM や記録媒体 (例えばハードディスクドライブ) に、上記各手段の機能を実現するためのプログラムを格納し、そのパーソナルコンピュータの CPU が RAM 等との間で適宜データを授受しながら、後述する一連の符号化処理を実行するようにすることができる。もちろん、音楽配信用の専用の音声符号化装置にすることで、より複雑な符号化モード設定や高音質の符号化処理を行うこともできる。

#### 【0029】

次に、本実施形態における音声符号化装置の動作とその音声符号化方法につい

て説明するとともに、データテーブル等の詳細構成について説明する。

#### 【0030】

まず、入力デジタルオーディオ信号が窓長分析手段110に入力されると、この窓長分析手段110により、入力信号が定常な信号の場合にはロングフレームの入力信号ブロックと、過渡的な信号の場合にはショートフレームの入力信号ブロックと分析される。また、入力デジタルオーディオ信号は、FFT分析手段100において、高速フーリエ変換され、この結果を用いて、聴覚心理分析手段130により人間の聴覚心理モデルに基づく、信号対マスク比(SMR)の値が算出される。この信号対マスク比(SMR)により、各スケールファクタバンドについて、マスキング効果によるマスキング閾値のレベルを超える範囲で、かつそのスケールファクタバンドのピークレベルまでの範囲が各スケールファクタバンドのスケールファクタとして特定される。また、最大スケールファクタバンド初期値設定手段140により、窓長分析手段110の分析結果および符号化モード情報指定手段120より得られた符号化ビットレートや入力信号のサンプリング周波数等の情報に応じて、最大スケールファクタバンド係数テーブル記憶手段180が参照され、最大スケールファクタバンド(max Sfb)初期値が設定される。

#### 【0031】

このように、最大スケールファクタバンド初期値設定手段140は、外部から指定された符号化モード情報指定手段120よりの情報に応じて、最大スケールファクタバンド係数テーブル記憶手段180を参照して、最大スケールファクタバンド(max Sfb)初期値および信号対マスク比(SMR)閾値を算出するが、本実施形態では、例えば、最大スケールファクタバンド(max Sfb)初期値を42、信号対マスク比(SMR)閾値を1.0とする。

#### 【0032】

次いで、最大スケールファクタバンド適応算出手段150により、上記最大スケールファクタバンド(max Sfb)の初期値および上述の信号対マスク比(SMR)の値を用いて、最大スケールファクタバンド(max Sfb)値が設定されると、この最大スケールファクタバンド(max Sfb)値および信号対マ



スケ比 (SMR) 等の情報を用いて、スペクトル処理手段 160 により MDCT 処理や TNS 処理等のスペクトル処理が行われ、次いで、量子化／符号化手段 170 によって量子化／符号化処理が実行される。本実施形態では、スケールファクタバンドごとに SMR 値を算出して符号化を行う時間周波数変換符号化方式を用いるが、本発明は、聴覚心理モデルを利用した各スケールファクタバンド信号に対するビット割当ての処理 (スケールファクタバンド信号ごとのスケールファクタに人間の聴覚特性に基づいた重み付けを行う処理) に注目したものではなく、周波数帯域順 (図 3 中では、最も低周波数側の Sfb「0」から最も高周波数側の Sfb「42」までの順) に設定されたスケールファクタバンドのうち、最大スケールファクタバンドの適応的な変化によって、符号化する周波数帯域を最適に保ち、無駄な符号化演算を削減する。

#### 【0033】

そのための最大スケールファクタバンド (max Sfb) 適応算出手段 150 における算出処理内容について、図 3 を参照して説明する。この図 3 は、複数のスケールファクタバンドについて、聴覚心理分析手段 130 により人間の聴覚心理モデルに基づいて算出された信号対マスク比 (SMR) 値と、その信号対マスク比 (SMR) についての閾値 (所定の閾値) とを示したグラフである。

#### 【0034】

最大スケールファクタバンド適応算出手段 150 は、最大スケールファクタバンド (max Sfb) 初期値における信号対マスク比 (SMR) 値と信号対マスク比 (SMR) の閾値とを比較し、信号対マスク比 (SMR) 値が信号対マスク比 (SMR) の閾値よりも小さい場合には、最大スケールファクタバンド (max Sfb) 値を 1 下げる。次いで、1 下げた最大スケールファクタバンド (max Sfb) における信号対マスク比 (SMR) 値と信号対マスク比 (SMR) 閾値とを比較し、信号対マスク比 (SMR) 値が信号対マスク比 (SMR) 閾値よりも小さいときには、最大スケールファクタバンド (max Sfb) 値を 1 下げる。

#### 【0035】

以上の動作を繰り返し行い、信号対マスク比 (SMR) 値が信号対マスク比 (

SMR) の閾値よりも大きくなった時のスケールファクタバンドに 1 を足し、その値を現フレームの最大スケールファクタバンド (max S f b) 値とする。

【 0 0 3 6 】

図 3 においては、信号対マスク比 (SMR) 値が信号対マスク比 (SMR) の閾値よりも大きくなった時のスケールファクタバンドが 3 8 であるため、最大スケールファクタバンド (max S f b) は 3 9 となる。

【 0 0 3 7 】

次に、上記最大スケールファクタバンド初期値テーブル 4 1 0 の構成について説明する。

図 4 および図 5 は最大スケールファクタバンド初期値テーブル 4 1 0 に格納されている情報の例を示す。この最大スケールファクタバンド初期値テーブル 4 1 0 は、最大スケールファクタバンド (max S f b) 初期値を記憶したものであって、上記符号化モード情報より入力される「ビットレート」、「サンプリング周波数」、「チャンネル数」、「窓長」を考慮して設定されている。図 4 (a) は「窓長」がロングフレーム、「チャンネル数」が 2 (ステレオ) の場合のテーブルであり、図 4 (b) は「窓長」がショートフレーム、「チャンネル数」が 2 (ステレオ) の場合のテーブルであり、図 5 (a) は「窓長」がロングフレーム、「チャンネル数」が 1 (モノラル) の場合のテーブルであり、図 5 (b) は「窓長」がショートフレーム、「チャンネル数」が 1 (モノラル) の場合のテーブルである。

【 0 0 3 8 】

最大スケールファクタバンド初期値テーブル 4 1 0 は、人間の耳に対して比較的感度の低い高帯域においては、符号化をし難くするように設定する。例えば、「ビットレート」が低くなり、使用可能なビットが少なくなる場合には、最大スケールファクタバンド (max S f b) 初期値を下げて、高帯域の符号化をし難くし、人間の耳に比較的感度の高い低周波数にビットを優先的に使用する。また「サンプリング周波数」が低くなった場合は、フレーム長が長くなり、使用可能ビットが多くなるため、高帯域の符号化をしやすいうように最大スケールファクタバンド (max S f b) 初期値を上げ、音質を向上させる。また、「チャンネル

数」が少ない場合には、1フレームにおける使用ビット数が少なくなるため、最大スケールファクタバンド (max Sfb) 初期値をあげ、音質を向上させる。また、「窓長」がショートフレームの場合は、入力音源が過渡的な音源であるため、高帯域のエネルギーが多くなるため、最大スケールファクタバンド (max Sfb) 初期値をあげるにより音質を向上させる。

## 【0039】

次に、上記信号対マスク比閾値テーブル（以下、SMR閾値テーブルという）420の構成について説明する。

図6および図7はSMR閾値テーブル420に格納されている情報の例を示している。このSMR閾値テーブル420は、最大スケールファクタバンド (max Sfb) の算出に必要な信号対マスク比 (SMR) の閾値を記憶したものであって、上記符号化モード情報のうち「ビットレート」、「サンプリング周波数」、「チャンネル数」、「窓長」を考慮してそれぞれの閾値が設定されている。ここで、図6(a)は「窓長」がロングフレーム、「チャンネル数」が2（ステレオ）の場合のテーブルであり、図6(b)は「窓長」がショートフレーム、「チャンネル数」が2（ステレオ）の場合のテーブルであり、図7(a)は「窓長」がロングフレーム、「チャンネル数」が1（モノラル）の場合のテーブルであり、図7(b)は「窓長」がショートフレーム、「チャンネル数」が1（モノラル）の場合のテーブルである。

## 【0040】

SMR閾値テーブル420は、最大スケールファクタバンド初期値テーブル410と同様に、人間の耳に対して比較的感度の低い高帯域においては、使用可能ビットが少ない場合に符号化をし難くするように設定されている。

## 【0041】

具体的には、例えば、「ビットレート」が低くなり、使用可能なビットが少なくなる場合には、信号対マスク比 (SMR) 閾値を上げ、高帯域の符号化をし難くし、人間の耳に比較的感度の高い低周波数にビットを優先的に使用する。また「サンプリング周波数」が低くなった場合は、フレーム長が長くなり、使用可能ビットが多くなるため、高帯域の符号化をしやすいように信号対マスク比 (SM

R) 閾値を下げて、音質を向上させる。また「チャンネル数」が少ない場合には、1フレームにおける使用ビット数が少なくなるため、信号対マスク比 (SMR) 閾値を下げて、音質を向上させる。また、「窓長」がショートフレームの場合は、入力音源が過渡的な音源であって高帯域のエネルギーが多くなるため、信号対マスク比 (SMR) 閾値を下げるにより音質を向上させる。

#### 【0042】

次に図8のフローチャートを参照して、本発明の音声符号化方法の一実施形態を上記装置のソフトウェアで実現する場合の処理の流れを説明する。

同図において、まず、入力デジタルオーディオ信号は、ステップS110において窓長分析手段110により、入力信号が定常な信号の場合にはロングフレームと分析され、過渡的な信号の場合にはショートフレームと分析される。

#### 【0043】

一方、入力デジタルオーディオ信号は、ステップS100においてFFT分析手段100によって、高速フーリエ変換 (FFT) され、この結果を用いて、聴覚心理分析手段130により人間の聴覚心理モデルに基づく、信号対マスク比 (SMR) 値が算出される (ステップS130)。

#### 【0044】

次いで、ステップS140において、最大スケールファクタバンド初期値設定手段140は、ステップS110における窓長分析手段110の分析結果および外部からの制御情報である符号化モード情報指定手段120より得られた (ステップS120) 符号化ビットレートや入力信号のサンプリング周波数等の情報より、最大スケールファクタバンド係数テーブル記憶手段180を参照し、最大スケールファクタバンド (max Sfb) の初期値および信号対マスク比 (SMR) の閾値を設定する。

#### 【0045】

次にステップS150において、最大スケールファクタバンド (max Sfb) 適応算出手段150は、最大スケールファクタバンド (max Sfb) 初期値における信号対マスク比 (SMR) 値と信号対マスク比 (SMR) 閾値とを比較し (ステップS151)、最大スケールファクタバンド (max Sfb) 初期値

における信号対マスク比 (SMR) 値が、信号対マスク比 (SMR) 閾値より小さい場合には、最大スケールファクタバンド (max Sfb) を 1 下げ (ステップ S152)、ステップ S151 に戻る。

【0046】

次ステップ S151 において、最大スケールファクタバンド (max Sfb) における信号対マスク比 (SMR) 値が、信号対マスク比 (SMR) 閾値より大きくなると、最大スケールファクタバンド (max Sfb) を 1 上げる (ステップ S153)。

次にステップ S160 において、スペクトル処理手段 160 は、MDCT 処理や TNS 処理等のスペクトル処理を行い、ステップ S170 において量子化／符号化手段 170 によって量子化／符号化が行われる。

【0047】

上記のように、本実施形態においは、入力信号に応じて最大スケールファクタバンドを適応的に算出することで、符号化する周波数帯域を常に好適にすることができる。したがって、人間の耳に聞こえ難く符号化に無駄となるスケールファクタバンドにおける演算量を削減することができ、符号量を有効に活用できる音声符号化方式を実現することができる。また、最大スケールファクタバンドは、符号化モード情報指定手段 120 から得られるビットレートおよびサンプリング周波数に応じて、最大スケールファクタバンド係数テーブル記憶手段 180 を参照することにより設定されるので、符号化モードに適した最大スケールファクタバンドの設定が可能となる。さらに、最大スケールファクタバンドの算出に符号化モード情報より得られるチャンネル数をも考慮するので、その符号化モードにより適した最大スケールファクタバンドの設定ができる。

【0048】

また、本実施形態においては、最大スケールファクタバンド適応算出手段 150 が、最大スケールファクタバンド係数テーブル記憶手段 180 を参照することで設定されるいずれか 1 つの最大スケールファクタバンドにおいて、前記聴覚心理分析手段 130 で算出された信号対マスク比 (SMR) と、窓長および符号化モード情報に応じて SMR 閾値テーブル 420 を参照することにより設定された

閾値とを比較し、算出された信号対マスク比 (SMR) が設定された SMR 閾値より小さい場合には最大スケールファクタバンド (S f b) を 1 ずつ下げていき、算出された信号対マスク比 (SMR) が設定された SMR 閾値より大きな値をとるスケールファクタバンド (S f b) になる 1 つ前のスケールファクタバンドを現在の符号化フレームの最大スケールファクタバンドとする。したがって、マスキング閾を常に好適に保つよう臨界帯域を適応的に制御し、無駄な符号化処理を省き、符号化効率を高めることができる。

【0049】

(実施の形態 2) 次に、本発明の実施の形態 2 にかかる音声符号化方式および装置について、図 9 ～図 12 を参照しながら説明する。

【0050】

図 9 は実施の形態 2 にかかる音声符号化装置のブロック図である。

【0051】

同図において、デジタルオーディオ信号である入力信号は、この入力信号を高速フーリエ変換 (FFT: Fast Fourier Transform) する FFT 分析手段 800 に入力され、スペクトル情報が得られる。この FFT 分析手段 800 の出力は聴覚心理分析手段 830 に入力され、そこで人間の聴覚心理モデルに基づき、信号対マスク比 (SMR) 値を算出する聴覚心理分析処理が実行される。

【0052】

入力信号は、また、窓長分析手段 810 にも入力され、入力信号が定常な信号の場合にはロングフレームの信号ブロックと、過渡的な信号の場合にはショートフレームの信号ブロックと分析される、さらに、入力信号のサンプリング周波数やビットレート等の符号化に必要な情報が符号化モード情報指定手段 820 で指定入力され、そこで設定された符号化モード情報が窓長分析手段 810 の出力と共に、最大スケールファクタバンド (max S f b) 初期値設定手段 840 に入力される。この最大スケールファクタバンド (max S f b) 初期値設定手段 840 は、入力情報を基に最大スケールファクタバンド (max S f b) 係数テーブル記憶手段 880 を参照し、最大スケールファクタバンド (max S f b) 初期値を設定する算出処理を行う。

## 【 0 0 5 3 】

また、FFT分析手段800からのスペクトル情報と、最大スケールファクタバンド (max Sfb) 初期値設定手段840からの初期値設定情報とに基づいて、最大スケールファクタバンド (max Sfb) 適応算出手段850が最大スケールファクタバンド (max Sfb) を入力信号の特性に応じて適応的に算出する。そして、算出された最大スケールファクタバンド (max Sfb) と聴覚心理分析手段830で算出された信号対マスク比 (SMR) 値に基づいて、スペクトル処理手段860によりMDCT処理やTNS処理等のスペクトル処理が実行され、処理量子化／符号化手段870により量子化／符号化される。ここで、最大スケールファクタバンド (max Sfb) 係数テーブル記憶手段880には、最大スケールファクタバンド値 (max Sfb) 算出に必要な初期値およびエネルギー閾値が記録されている。

## 【 0 0 5 4 】

次に本実施の形態における音声符号化装置の動作について説明する。

まず、入力デジタルオーディオ信号が、窓長分析手段810により入力信号が定常な信号の場合にはロングフレーム、過渡的な信号の場合にはショートフレームと分析される。

一方、入力デジタルオーディオ信号が、FFT分析手段800において、高速フーリエ変換 (FFT: Fast Fourier Transform) され、この結果を用いて、聴覚心理分析手段830により人間の聴覚心理モデルに基づく、信号対マスク比 (SMR) 値が算出される。

## 【 0 0 5 5 】

次に、最大スケールファクタバンド (max Sfb) 初期値設定手段840により、窓長分析手段810の分析結果および符号化モード情報指定手段820からの符号化ビットレートや入力信号のサンプリング周波数等の符号化モード情報 (外部からの制御情報) に応じて、最大スケールファクタバンド (max Sfb) 係数テーブル記憶手段880の対応するデータが参照され、最大スケールファクタバンド (max Sfb) の初期値が設定される。

## 【 0 0 5 6 】

さらに、最大スケールファクタバンド (max Sfb) 適応算出手段 850 により、前記最大スケールファクタバンド (max Sfb) 初期値および FFT 分析手段 800 において算出されたスペクトル情報から、最大スケールファクタバンド (max Sfb) 値を設定する算出処理がなされる。

## 【0057】

次いで、この最大スケールファクタバンド (max Sfb) 値および信号対マスク比 (SMR) 等の情報を用いて、スペクトル処理手段 860 により MDCT 処理や TNS 処理等のスペクトル処理が行われ、その出力に基づいて処理量子化／符号化手段 870 で量子化／符号化の処理がなされる。

## 【0058】

上記最大スケールファクタバンド (max Sfb) の初期値設定では、外部からの制御情報である符号化モード情報指定手段 820 からの指定情報に応じて、最大スケールファクタバンド (max Sfb) 係数テーブル記憶手段 880 を参照することにより、最大スケールファクタバンド (max Sfb) 初期値およびエネルギー閾値が算出されるが、本実施の形態では、各スケールファクタバンド内のエネルギー値を算出し、これを予め設定したエネルギー閾値と比較するようになっている。

## 【0059】

具体的には、図 10 に示すように、最大スケールファクタバンド (max Sfb) の初期値を 42 とし、エネルギー閾値を 100、000 としている。次に、最大スケールファクタバンド (max Sfb) 適応算出手段 850 の算出処理について説明する。まず、上記 FFT 分析手段 800 より算出されたスペクトル  $\text{spectral}[i]$  から、下記の数式 1 に示すように、各サブバンドのエネルギー値  $\text{energy}[\text{sfb}]$  を算出する。



【数 1】

$$energy[sfb] = \sum_{sfb=0}^{sfb=\max Sfb} \sum_{start[sfb]}^{end[sfb]} spectral[i] \cdot spectral[i]$$

ここで、

sfb：スケールファクタバンド

maxSfb：最大スケールファクタバンド(maxSfb) 初期値

start[sfb]：スケールファクタバンドの開始点

end[sfb]：スケールファクタバンドの終了点  
である。

【0 0 6 0】

次に、図 1 0 を参照して、最大スケールファクタバンド (max Sfb) 適応算出手段 8 5 0 の動作について説明する。

【0 0 6 1】

図 1 0 は、各スケールファクタバンドのエネルギー値および上述のエネルギー閾値を示したものである。最大スケールファクタバンド (max Sfb) 適応算出手段 1 5 0 は、最大スケールファクタバンド (max Sfb) 初期値におけるエネルギー値とエネルギー閾値とを比較し、スケールファクタバンドのエネルギー値がエネルギー閾値よりも小さいときには、最大スケールファクタバンド (max Sfb) 値を 1 下げる。次に、1 下げた最大スケールファクタバンド (max Sfb) におけるエネルギー値とエネルギー閾値とを比較し、スケールファクタバンドのエネルギー値がエネルギー閾値よりも小さいときには、最大スケールファクタバンド (max Sfb) 値を 1 下げる。

【0 0 6 2】

以上の動作を繰り返し行い、スケールファクタバンドのエネルギー値がエネルギー閾値よりも大きくなった時のスケールファクタバンドに 1 足した値を現フレームの最大スケールファクタバンド (max Sfb) 値とする。

【0 0 6 3】

図 1 0 においては、スケールファクタバンドのエネルギー値がエネルギー閾値よりも大きくなった時のスケールファクタバンドが 3 8 であるため、最大スケールファクタバンド (max Sfb) は 3 9 となる。

## 【 0 0 6 4 】

最大スケールファクタバンド (max Sfb) 係数テーブル記憶手段 880 は、図 2 に示したものと同様に最大スケールファクタバンド初期値テーブル 410 および信号対マスク比 (SMR) 値テーブル 420 で構成される。ただし、この場合の最大スケールファクタバンド初期値テーブル 410 の構成が図 4 および図 5 に示したものと同様であるのに対して、本実施形態における閾値テーブルは、図 11 および図 12 に示すように、最大スケールファクタバンド (max Sfb) 適応算出手段 850 に必要なスケールファクタバンドごとのエネルギー閾値を記憶したものであって、上記符号化モード情報より入力される「ビットレート」、「サンプリング周波数」、「チャンネル数」、「窓長」を考慮して設定されている。この場合、図 11 (a) は「窓長」がロングフレーム、「チャンネル数」が 2 (ステレオ) の場合のテーブルであり、図 11 (b) は「窓長」がショートフレーム、「チャンネル数」が 2 (ステレオ) の場合のテーブルであり、図 12 (a) は「窓長」がロングフレーム、「チャンネル数」が 1 (モノラル) の場合のテーブルであり、図 12 (b) は「窓長」がショートフレーム、「チャンネル数」が 1 (モノラル) の場合のテーブルである。

## 【 0 0 6 5 】

この図 11 および図 12 に示す閾値テーブル 420 E は、最大スケールファクタバンド初期値テーブル 410 と同様に、人間の耳に対して比較的感度の低い高帯域においては、使用可能ビットが少ない場合には、符号化をし難くするように設定する。例えば、「ビットレート」が低くなり、使用可能なビットが少なくなる場合には、エネルギー閾値を上げ、高帯域の符号化をし難くし、人間の耳に比較的感度の高い低周波数にビットを優先的に使用する。また「サンプリング周波数」が低くなった場合は、フレーム長が長くなり、使用可能ビットが多くなるため、高帯域の符号化をしやすいようにエネルギー閾値を下げて、音質を向上させるようになっている。また「チャンネル数」が少ない場合には、1 フレームにおける使用ビット数が少なくなるため、エネルギー閾値を下げ、音質を向上させる。さらに、「窓長」がショートフレームの場合は、入力音源が過渡的な音源であって、高帯域のエネルギーが多くなるため、エネルギー閾値を下げることでより

音質を向上させる。

【0066】

次に、図16のフローチャートを参照して、本発明の音声符号化方法の他の実施形態を上記装置のソフトウェアで実現する場合の処理の流れを説明する。

【0067】

同図に示すように、まず、入力デジタルオーディオ信号は、ステップS810において窓長分析手段810により、入力信号が定常な信号の場合にはロングフレーム、過渡的な信号の場合にはショートフレームと分析される。

【0068】

一方、入力デジタルオーディオ信号は、ステップS800でFFT分析手段800によって高速フーリエ変換（FFT：Fast Fourier Transform）され、この結果を用いて、聴覚心理分析手段830により人間の聴覚心理モデルに基づく、信号対マスク比（SMR）値が算出される（ステップS830）。

【0069】

次のステップS840においては、最大スケールファクタバンド（max Sfb）初期値設定手段840により、ステップS810における窓長分析手段810の分析結果および符号化モード情報指定手段820から得られた（ステップS820）符号化ビットレートや入力信号のサンプリング周波数等より、最大スケールファクタバンド（max Sfb）係数テーブル記憶手段880を参照し、最大スケールファクタバンド（max Sfb）初期値およびエネルギー閾値を設定する。

【0070】

次いで、ステップS850において、最大スケールファクタバンド（max Sfb）適応算出手段850により、FFT分析手段800で算出されたスペクトル情報から各スケールファクタバンド内のエネルギー値が算出される（ステップS851）。

【0071】

次いで、ステップS852において、最大スケールファクタバンド（max Sfb）初期値におけるエネルギー値とエネルギー閾値とが比較され、スケールフ

ファクタバンドのエネルギー値がエネルギー閾値よりも小さいときには、最大スケールファクタバンド (max Sfb) 値を 1 下げて (ステップ S853)、ステップ S851 に戻る処理がされる。ステップ S851 で、最大スケールファクタバンド (max Sfb) におけるエネルギー値が、エネルギー閾値より大きくなると、最大スケールファクタバンド (max Sfb) を 1 上げる処理 (ステップ S854) がされる。

## 【0072】

次いで、ステップ S860 に進み、スペクトル処理手段 860 により、MDC T 処理や TNS 処理等のスペクトル処理が実行された後、ステップ S870 において、処理量子化／符号化手段 870 によって量子化処理がされ、一連の符号化の処理が終了する。

## 【0073】

上記のように、本実施形態においても、実施の形態 1 と同様に、入力信号に応じて最大スケールファクタバンドを適応的に算出することで、符号化する周波数帯域を常に好適にすることができ、演算量の削減と符号量の有効活用を図ることができ、上述と同様な効果を得ることができる。

## 【0074】

また、本実施形態においては、最大スケールファクタバンド適応算出手段 850 が、最大スケールファクタバンド係数テーブル記憶手段 880 を参照することにより設定されるいずれか 1 つの最大スケールファクタバンドにおいて入力信号のスペクトルのエネルギー値を算出した後、その算出値と前記窓長および前記符号化モード情報に応じて記閾値テーブルを参照することにより設定されたエネルギー閾値とを比較して、最適な最大スケールファクタバンドに適応的に制御するので、マスキング閾および臨界帯域を常に好適に保って無駄な符号化処理を省き、符号化効率を高めることができる。

## 【0075】

(実施の形態 3) 次に、本発明の実施の形態 3 における音声符号化方式および装置について、図を参照しながら説明する。

図 13 は実施の形態 1 における音声符号化装置における最大スケールファクタ

バンド適応算出手段のブロック図を示したものである。

【0076】

図13に示す音声符号化装置は、入力デジタルオーディオ信号を高速フーリエ変換（FFT：Fast Fourier Transform）するFFT分析手段1100と、入力信号が定常な信号の場合にはロングフレーム、過渡的な信号の場合にはショートフレームと分析する窓長分析手段1110と、入力信号のサンプリング周波数やビットレート等の符号化に必要な情報を設定するための符号化モード情報指定手段1120と、人間の聴覚心理モデルに基づいて信号対マスク比（SMR）値を算出する聴覚心理分析手段1130と、最大スケールファクタバンド（max Sfb）初期値を算出する最大スケールファクタバンド（max Sfb）初期値設定手段1140と、最大スケールファクタバンド（max Sfb）を入力信号の特性に応じて適応的に算出する最大スケールファクタバンド（max Sfb）適応算出手段1150と、MDCT処理やTNS処理等のスペクトル処理を行うスペクトル処理手段1160と、量子化および符号化処理を行う処理量子化／符号化手段1170と、最大スケールファクタバンド値（max Sfb）算出に必要な初期値および閾値を記憶している最大スケールファクタバンド（max Sfb）係数テーブル記憶手段1180とを備えている。

【0077】

図15に示すように、最大スケールファクタバンド（max Sfb）係数テーブル1180には、最大スケールファクタバンド（max Sfb）初期値テーブル1310および信号対マスク比（SMR）閾値テーブル1320に加えて、最小スケールファクタバンド（min Sfb）テーブル1330が設けられている。この最大スケールファクタバンド（max Sfb）初期値テーブル1310の構成としては、例えば、図4および図5に示したような構成が用いられる。また、上記信号対マスク比（SMR）閾値テーブル1320の構成としては、例えば、図6および図7に示したような構成が用いられる。前記最小スケールファクタバンド（min Sfb）テーブル1330は、符号化モード情報より入力される「ビットレート」、「サンプリング周波数」、「チャンネル数」、「窓長」を考慮して設定されており、必ず符号化処理を行う最大スケールファクタバンドの値が

設定されている（例えば 3 k H z 以下は常に符号化処理を行うように設定する）。

#### 【0078】

次に、本実施の形態における音声符号化装置の動作について説明すると、まず、入力デジタルオーディオ信号が、窓長分析手段 1 1 1 0 により入力信号が定常な信号の場合にはロングフレーム、過渡的な信号の場合にはショートフレームと分析される。一方、入力デジタルオーディオ信号は、FFT 分析手段 1 1 0 0 において高速フーリエ変換（FFT : Fast Fourier Transform）され、この結果を用いて、聴覚心理分析手段 1 1 3 0 により人間の聴覚心理モデルに基づく、信号対マスク比（SMR）値が算出される。次に、最大スケールファクタバンド（max S f b）初期値設定手段 1 1 4 0 により、窓長分析手段 1 1 1 0 の分析結果と、符号化モード情報指定手段 1 1 2 0 から得られた符号化ビットレートや入力信号のサンプリング周波数等の符号化モード情報とに応じて、最大スケールファクタバンド（max S f b）係数テーブル記憶手段 1 1 8 0 が参照され、最大スケールファクタバンド（max S f b）初期値が設定される。

#### 【0079】

次いで、最大スケールファクタバンド（max S f b）適応算出手段 1 1 5 0 により、上記最大スケールファクタバンド（max S f b）初期値および上述の信号対マスク比（SMR）値および上記符号化モード情報指定手段 1 1 2 0 より得られた窓長の情報を用いて、最大スケールファクタバンド（max S f b）値を設定する算出処理がされる。

#### 【0080】

そして、この最大スケールファクタバンド（max S f b）値および信号対マスク比（SMR）等の情報を用いて、スペクトル処理手段 1 1 6 0 によって、M D C T 処理や T N S 処理等のスペクトル処理が行われた後、処理量子化／符号化手段 1 1 7 0 によって量子化／符号化の処理が実行される。本実施形態においては、最大スケールファクタバンド（max S f b）初期値設定手段 1 1 4 0 は、外部からの制御情報である符号化モード情報指定手段 1 1 2 0 のへの指定入力情報に応じて、最大スケールファクタバンド（max S f b）係数テーブル 1

180を参照することより、最大スケールファクタバンド (max Sfb) 初期値および信号対マスク比 (SMR) 閾値を算出する。具体的には、図14に示すように、本実施の形態では、最大スケールファクタバンド (max Sfb) 初期値を13、信号対マスク比 (SMR) 閾値を1.0としている。

## 【0081】

次に、図14を参照しながら、上記最大スケールファクタバンド (max Sfb) 適応算出手段1150での算出処理について説明する。なお、同図は、聴覚心理分析手段1130で人間の聴覚心理モデルに基づき算出された信号対マスク比 (SMR) の値とその信号対マスク比 (SMR) の閾値の関係を示している。

## 【0082】

まず、最大スケールファクタバンド (max Sfb) 適応算出手段1150は、最大スケールファクタバンド (max Sfb) 初期値における信号対マスク比 (SMR) 値と信号対マスク比 (SMR) 閾値とを比較し、信号対マスク比 (SMR) 値が信号対マスク比 (SMR) 閾値よりも小さいときには、最大スケールファクタバンド (max Sfb) 値を1下げる。次に、1下げた最大スケールファクタバンド (max Sfb) における信号対マスク比 (SMR) 値と信号対マスク比 (SMR) 閾値とを比較し、信号対マスク比 (SMR) 値が信号対マスク比 (SMR) 閾値よりも小さいときには、最大スケールファクタバンド (max Sfb) 値を1下げる。そして、以上の動作を繰り返し行い、信号対マスク比 (SMR) 値が信号対マスク比 (SMR) 閾値よりも大きくなった時のスケールファクタバンドに1足した値を現フレームの最大スケールファクタバンド (max Sfb) 値とする。

## 【0083】

ここで、現フレームの最大スケールファクタバンド (max Sfb) 値が前記最小スケールファクタバンド (min Sfb) 以下である場合には、最小スケールファクタバンド (min Sfb) に1足した値を現フレームの最大スケールファクタバンド (max Sfb) 値とする。本実施形態においては、上記最小スケールファクタバンド (min Sfb) を11とすると、図14においては、信号対マスク比 (SMR) 値が信号対マスク比 (SMR) 閾値よりも大きくなった

時のスケールファクタバンドが6であるため、最大スケールファクタバンド ( $\max Sfb$ ) は7となる。このとき、最大スケールファクタバンド ( $\max Sfb$ ) は、最小スケールファクタバンド ( $\min Sfb$ ) よりも小さいため、最大スケールファクタバンドをこの最小スケールファクタバンドを基に例えば ( $\min Sfb + 1$  として) 11とする。これにより、符号化する帯域が狭いときには、より広い帯域の符号化を行うことにより音質を向上させることができる。

## 【0084】

次に図17のフローチャートを参照して、本発明の音声符号化方法を上記装置のソフトウェアで実現する場合の処理の流れを説明する。

まず、入力デジタルオーディオ信号が、ステップS1110において窓長分析手段1110により、入力信号が定常な信号の場合にはロングフレーム、過渡的な信号の場合にはショートフレームと分析される。

## 【0085】

一方、入力デジタルオーディオ信号は、ステップS1100においてFFT分析手段1100によって、高速フーリエ変換 (FFT: Fast Fourier Transform) される。そして、この結果を用いて、聴覚心理分析手段1130により、人間の聴覚心理モデルに基づく、信号対マスク比 (SMR) 値が算出される (ステップS1130)。

## 【0086】

次のステップS1140においては、最大スケールファクタバンド ( $\max Sfb$ ) 初期値設定手段1140により、ステップS1110における窓長分析手段1110の分析結果および符号化モード情報指定手段1120より得られた (ステップS1120) 符号化ビットレートや入力信号のサンプリング周波数等の符号化モード情報に応じて、最大スケールファクタバンド係数テーブル記憶手段1180が参照され、最大スケールファクタバンド ( $\max Sfb$ ) 初期値および信号対マスク比 (SMR) の閾値がそれぞれ設定される。

## 【0087】

次いで、ステップS1150において、最大スケールファクタバンド ( $\max Sfb$ ) 適応算出手段1150により、最大スケールファクタバンド ( $\max Sfb$ )



f b) 初期値における信号対マスク比 (SMR) 値と信号対マスク比 (SMR) 閾値とが比較され (ステップ S 1 1 5 1)、最大スケールファクタバンド (max S f b) 初期値における信号対マスク比 (SMR) 値が、信号対マスク比 (SMR) 閾値より小さい場合には、最大スケールファクタバンド (max S f b) を 1 下げて (ステップ S 1 1 5 2) ステップ S 1 1 5 1 に戻る処理がされる。ステップ S 1 1 5 1 で、最大スケールファクタバンド (max S f b) における信号対マスク比 (SMR) 値が信号対マスク比 (SMR) 閾値より大きくなると、最大スケールファクタバンド (max S f b) を 1 上げる処理 (ステップ S 1 1 5 3) がなされる。

## 【0088】

次に、ステップ S 1 1 5 4 において、最大スケールファクタバンド (max S f b) と最小スケールファクタバンド (min S f b) が比較され (ステップ S 1 1 5 4)、最大スケールファクタバンド (max S f b) が最小スケールファクタバンド (min S f b) より小さい場合には、最大スケールファクタバンド (max S f b) を最小スケールファクタバンド (min S f b) に 1 を足した値にする (ステップ S 1 1 5 5) 処理がなされる。次いで、ステップ S 1 1 6 0 において、スペクトル処理手段 1 1 6 0 により、MDC T 処理や T N S 処理等のスペクトル処理がされた後、ステップ S 1 1 7 0 で処理量子化／符号化手段 1 1 7 0 によって量子化／符号化が行われる。上記のように、本実施形態においても、実施の形態 1、2 と同様に、入力信号に応じて最大スケールファクタバンドを適応的に算出することで、符号化する周波数帯域を常に好適にすることができ、演算量の削減と符号量の有効活用を図ることができ、上述と同様な効果を得ることができる。

## 【0089】

また、本実施形態においては、窓長分析手段 1 1 1 0 により入力信号が過渡的であることが分析され、ビットレートが所定値より高く、スケールファクタバンドのそれぞれに割付け可能なビット数が所定値を超えて十分にある状態においては、最大スケールファクタバンド適応算出手段 1 1 5 0 が算出した最大スケールファクタバンドが所定の最小スケールファクタバンドより低い場合には、最大ス

ケールファクタバンドの設定値を算出した値より所定の増分だけ上げるようにしているので、削減した周波数帯域の符号量（ビット割付量）を、音質に影響する周波数帯域の符号化に有効活用することができ、復号化時の音質を向上させることができる。

## 【 0 0 9 0 】

以上に述べた各実施の形態においては、本発明の音声符号化装置の各機能又は音声符号化方法の各ステップを実現するためのソフトウェアプログラムを準備して、これをハードディスクその他の記録媒体に記録したコンピュータ等により実行することとしていたが、本発明は、そのようなプログラムをコンピュータ読取可能な他の記録媒体、例えば磁気ディスク（フロッピーディスク、ハードディスク等）、光ディスク（CD-ROM、DVD等）、半導体メモリなどに記録したものであってもよく、これを読取って上記プログラムを実行するハードウェア資源としては、パーソナルコンピュータ程度から音楽配信用の専用音声符号化システムまで各種のものが採用できることはいうまでもない。

## 【 0 0 9 1 】

## 【発明の効果】

本発明の音声符号化方式によれば、人間の耳には聞こえない高域における符号量を削減することができ、その削減したビットをエネルギーの高い帯域に割り振ることにより音質を向上させることができる。また、高域の量子化／符号化における演算を削減することができるため、演算量を少なくすることができる。さらに、符号化する帯域が狭いときには、より広い帯域の符号化を行うことで、音質を向上させることができる。

## 【図面の簡単な説明】

## 【図 1】

本発明の実施の形態 1 における音声符号化装置のブロック図

## 【図 2】

本発明の実施の形態 1 における音声符号化装置で使用する最大スケールファクタバンド（max Sfb）係数テーブルの構成を示す図

## 【図 3】

本発明の実施の形態 1 における音声符号化装置で使用する最大スケールファクタバンド (max Sfb) 適応算出手段の処理の一例を示した模式図

【図 4】

本発明の実施の形態 1 における音声符号化装置で使用する最大スケールファクタバンド (max Sfb) 初期値テーブルの構成を示す図

【図 5】

図 4 に続く、上記実施の形態 1 における音声符号化装置で使用する最大スケールファクタバンド (max Sfb) 初期値テーブルの構成を示す図

【図 6】

本発明の実施の形態 1 における音声符号化装置で使用する信号対マスク比 (SMR) 閾値テーブルの構成を示す図

【図 7】

図 6 に続く、上記実施の形態 1 における音声符号化装置で使用する信号対マスク比 (SMR) 閾値テーブルの構成を示す図

【図 8】

本発明の実施の形態 1 における音声符号化方式を示すフローチャート

【図 9】

本発明の実施の形態 2 における音声符号化装置のブロック図

【図 10】

本発明の実施の形態 2 における音声符号化装置で使用する最大スケールファクタバンド (max Sfb) 適応算出手段の処理の一例を示した模式図

【図 11】

本発明の実施の形態 2 における音声符号化装置で使用する閾値テーブルの構成を示す図

【図 12】

図 11 に続く、上記実施の形態 2 における音声符号化装置で使用する閾値テーブルの構成を示す図

【図 13】

本発明の実施の形態 3 における音声符号化装置のブロック図

【図 1 4】

本発明の実施の形態 3 における音声符号化装置で使用する最大スケールファクタバンド (max Sfb) 適応算出手段の処理の一例を示した模式図

【図 1 5】

本発明の実施の形態 3 における音声符号化装置で使用する最大スケールファクタバンド (max Sfb) 係数テーブルの構成を示す図

【図 1 6】

本発明の実施の形態 2 における音声符号化方式を示すフローチャート

【図 1 7】

本発明の実施の形態 3 における音声符号化方式を示すフローチャート

【図 1 8】

従来例の最大スケールファクタバンド値 (max Sfb) を示すテーブル図

【図 1 9】

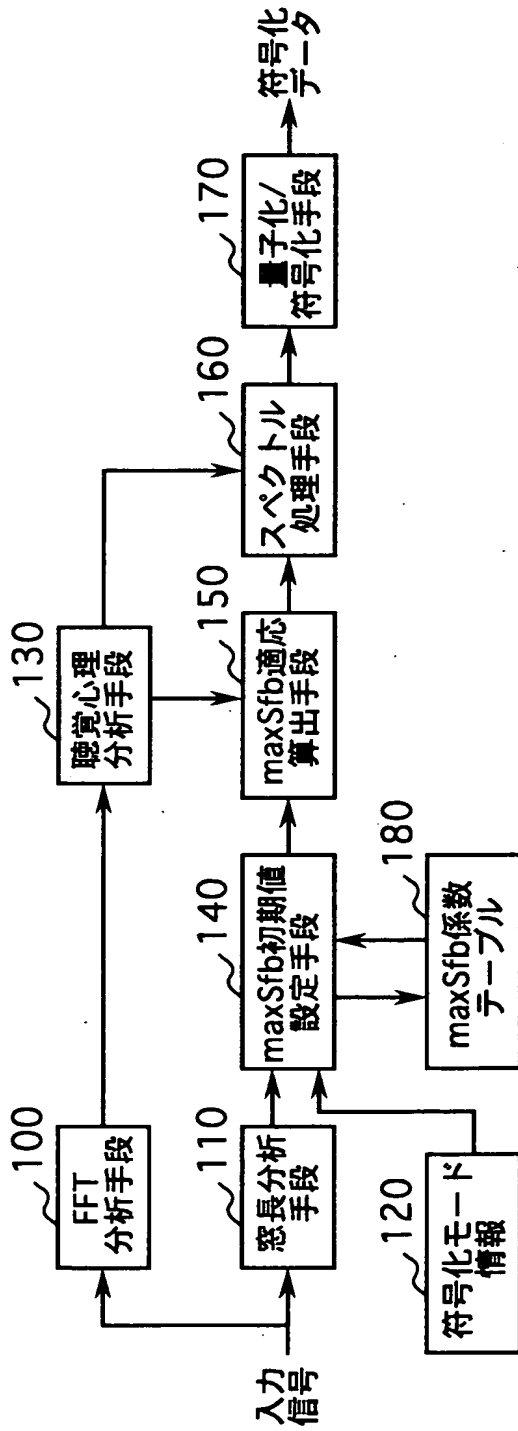
従来例の音声符号装置のブロック図

【符号の説明】

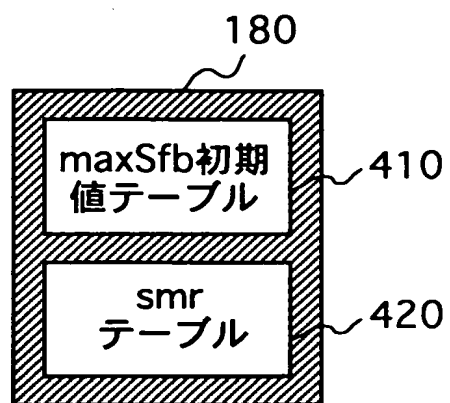
1 0 0、8 0 0、1 1 0 0	F F T 分析手段
1 1 0、8 1 0、1 1 1 0	窓長分析手段
1 2 0、8 2 0、1 1 2 0	符号化モード情報指定手段
1 3 0、8 3 0、1 1 3 0	聴覚心理分析手段
1 4 0、8 4 0、1 1 4 0	最大スケールファクタバンド初期値設定手段
1 5 0、8 5 0、1 1 5 0	最大スケールファクタバンド適応算出手段
1 6 0、8 6 0、1 1 6 0	スペクトル処理手段
1 7 0、8 7 0、1 1 7 0	量子化／符号化手段
1 8 0、8 8 0、1 1 8 0	最大スケールファクタバンド係数テーブル記憶手段 (閾値テーブル記憶手段)
4 1 0、1 3 1 0	最大スケールファクタバンド初期値テーブル
4 2 0、4 2 0 E、1 3 2 0	信号対マスク比 (SMR) 閾値テーブル
1 3 3 0	最小スケールファクタバンド (min Sfb) テーブル

【書類名】 図面

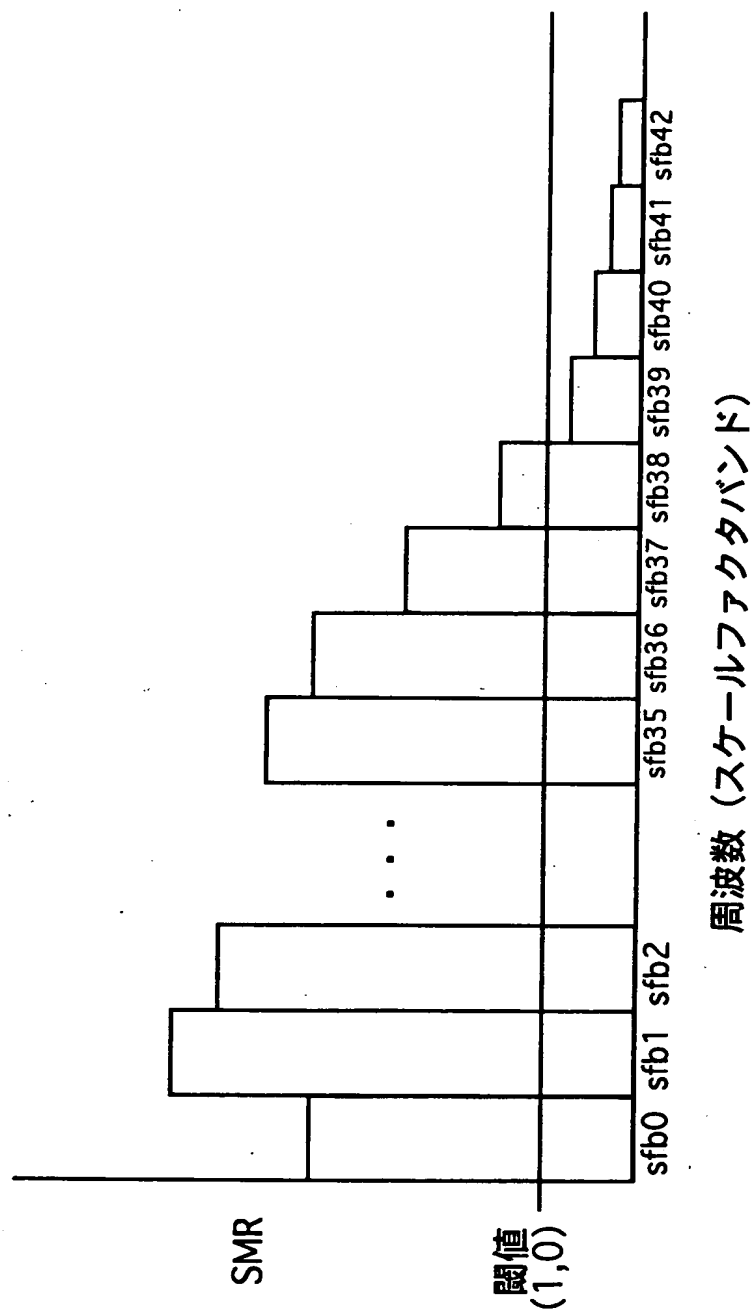
【図 1】



【図 2】



【図 3】



【図 4】

410

周波数 ヒット	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	49	49	51	47	47	43	43	43	40
96000	43	44	48	47	47	43	43	43	40
80000	44	43	47	47	47	43	43	43	40
64000	37	38	40	45	47	43	43	43	40

(a)

周波数 ヒット	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	13	14	14	15	15	15	15	15	15
96000	11	13	13	15	15	15	15	15	15
80000	10	12	12	15	15	15	15	15	15
64000	9	11	11	15	15	15	15	15	15

(b)



【図 5】

410

(a)

周波数 ピット	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	49	49	51	47	47	43	43	43	40
96000	45	45	51	47	47	43	43	43	40
80000	44	44	51	47	47	43	43	43	40
64000	39	38	40	45	47	43	43	43	40

(b)

周波数 ピット	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	13	14	14	15	15	15	15	15	15
96000	12	13	14	15	15	15	15	15	15
80000	11	12	14	15	15	15	15	15	15
64000	10	11	13	15	15	15	15	15	15

【図 6】

420

(a)

周波数 ビットレート	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	0.1	0.1	0.1	0.1	0.05	0.05	0.05	0.05	0.05
96000	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
80000	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
64000	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

(b)

周波数 ビットレート	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	1.0	1.0	1.0	0.1	0.1	0.1	0.1	0.1	0.1
96000	1.0	1.0	1.0	1.0	1.0	0.1	0.1	0.1	0.1
80000	1.0	1.0	1.0	1.0	1.0	0.1	0.1	0.1	0.1
64000	1.0	1.0	1.0	1.0	1.0	1.0	0.1	0.1	0.1

【図 7】

420

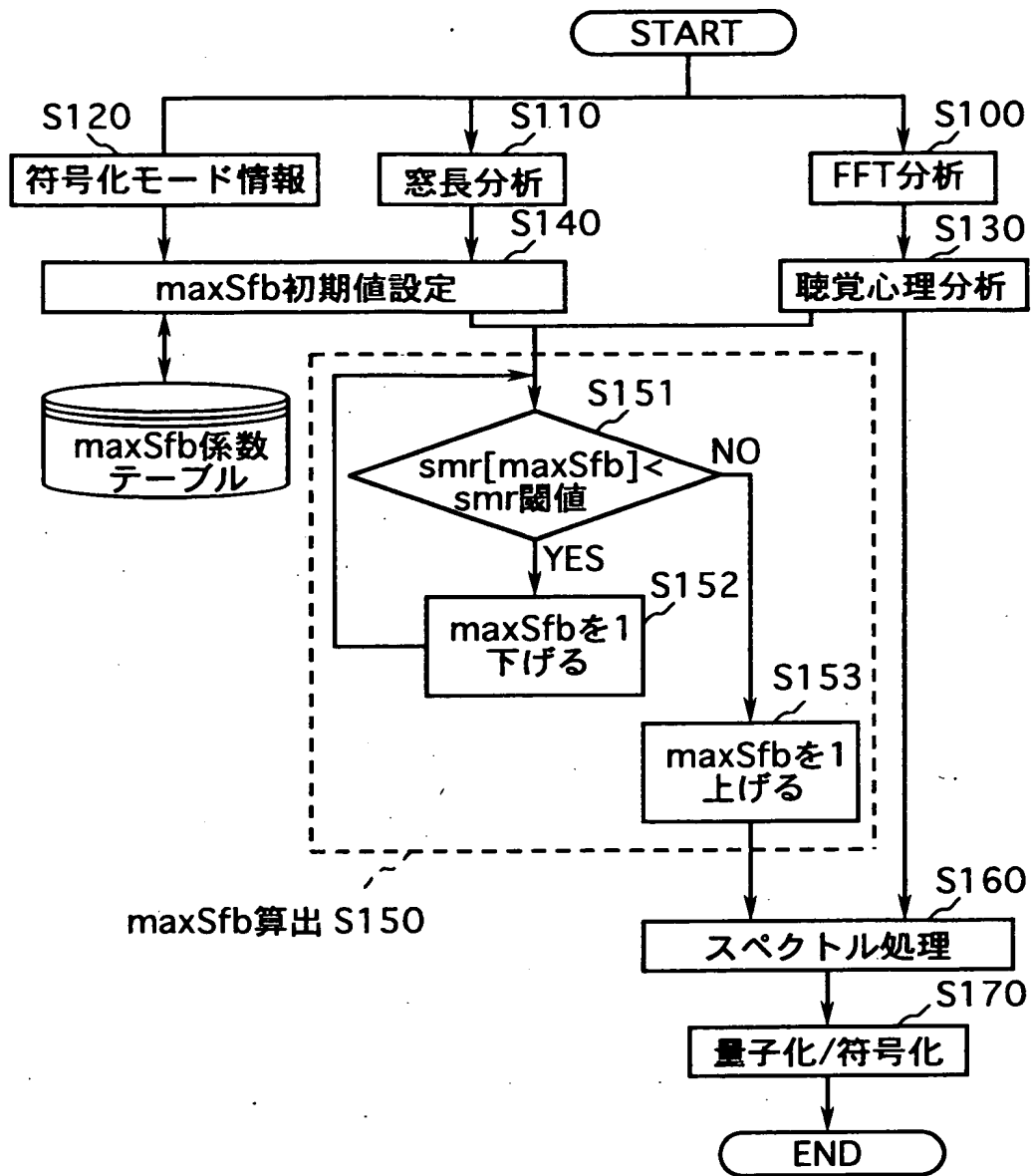
(a)

周波数 ヒットレ-ト	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	0.1	0.1	0.1	0.05	0.05	0.05	0.05	0.05	0.05
96000	0.1	0.1	0.1	0.1	0.1	0.05	0.05	0.05	0.05
80000	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
64000	0.5	0.5	0.5	0.5	0.5	0.5	0.1	0.1	0.1

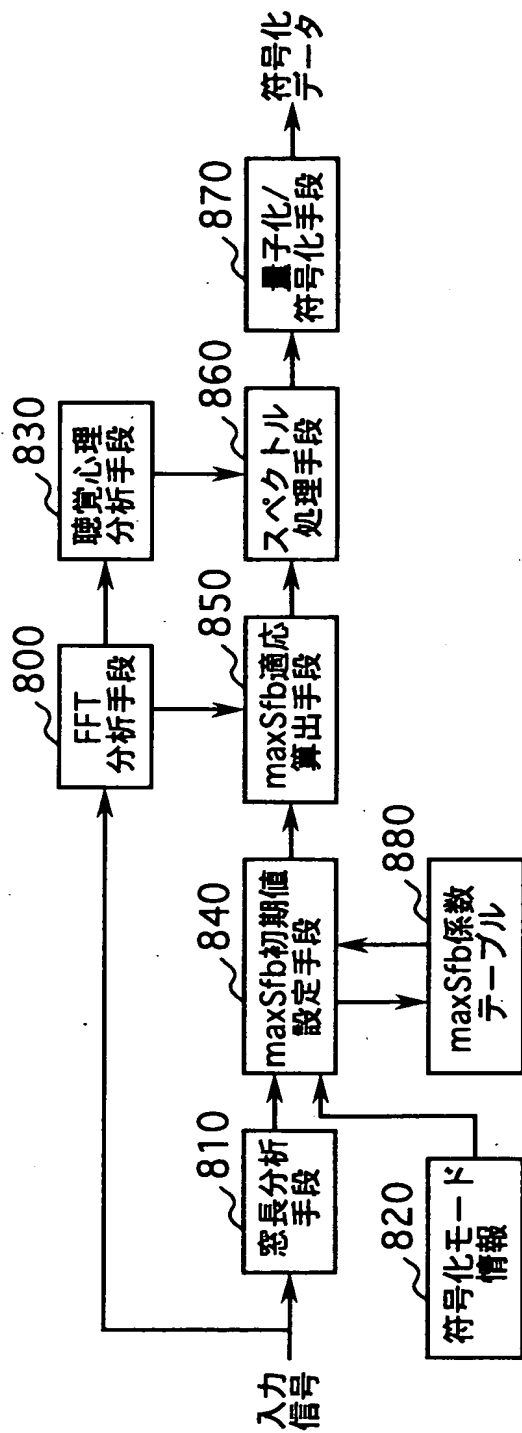
(b)

周波数 ヒットレ-ト	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	1.0	1.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1
96000	1.0	1.0	1.0	0.1	0.1	0.1	0.1	0.1	0.1
80000	1.0	1.0	1.0	0.1	0.1	0.1	0.1	0.1	0.1
64000	1.0	1.0	1.0	1.0	1.0	1.0	0.1	0.1	0.1

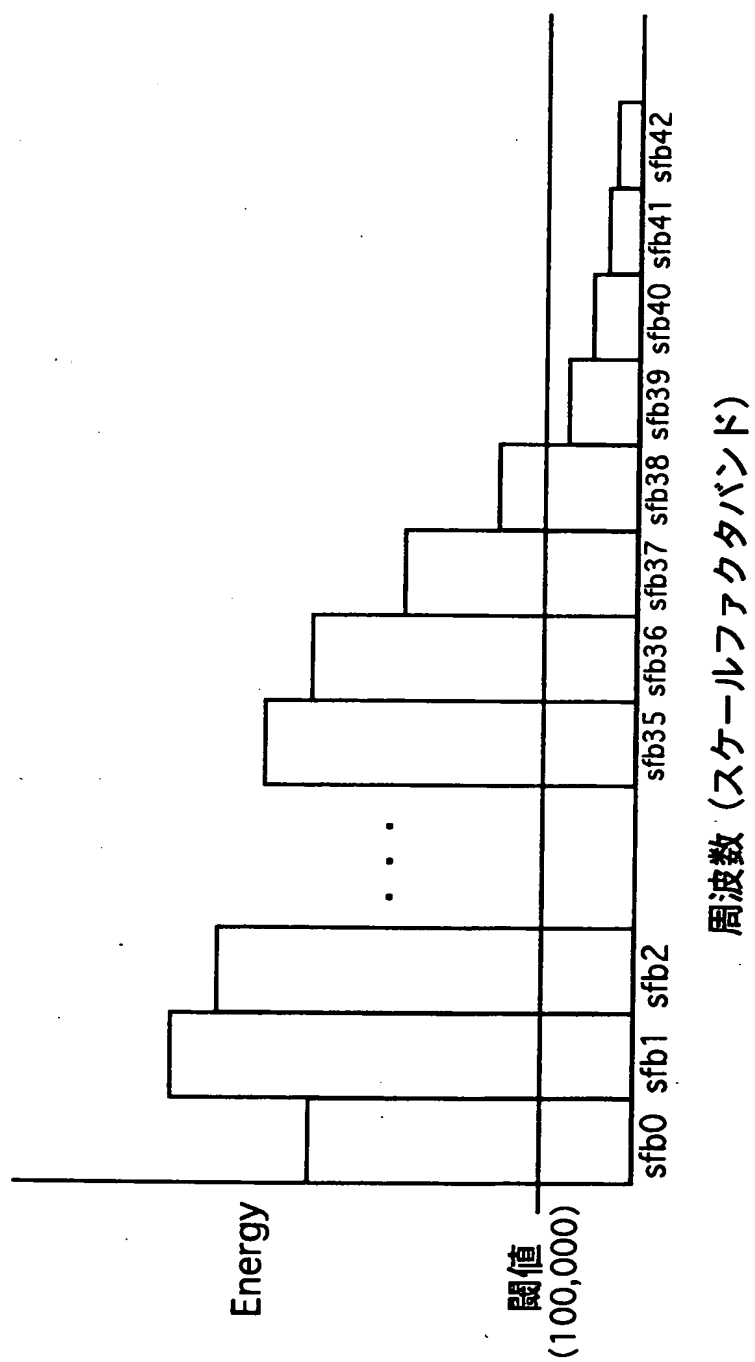
【図 8】



【図9】



【図 10】



【図 1 1】

420E

(a)

周波数 ビット	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	100000	100000	100000	100000	50000	50000	50000	50000	50000
96000	100000	100000	100000	100000	100000	100000	100000	100000	100000
80000	100000	100000	100000	100000	100000	100000	100000	100000	100000
64000	500000	500000	500000	500000	500000	500000	500000	500000	500000

(b)

周波数 ビット	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000
96000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000
80000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000
64000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000

【図 1 2】

420E

(a)

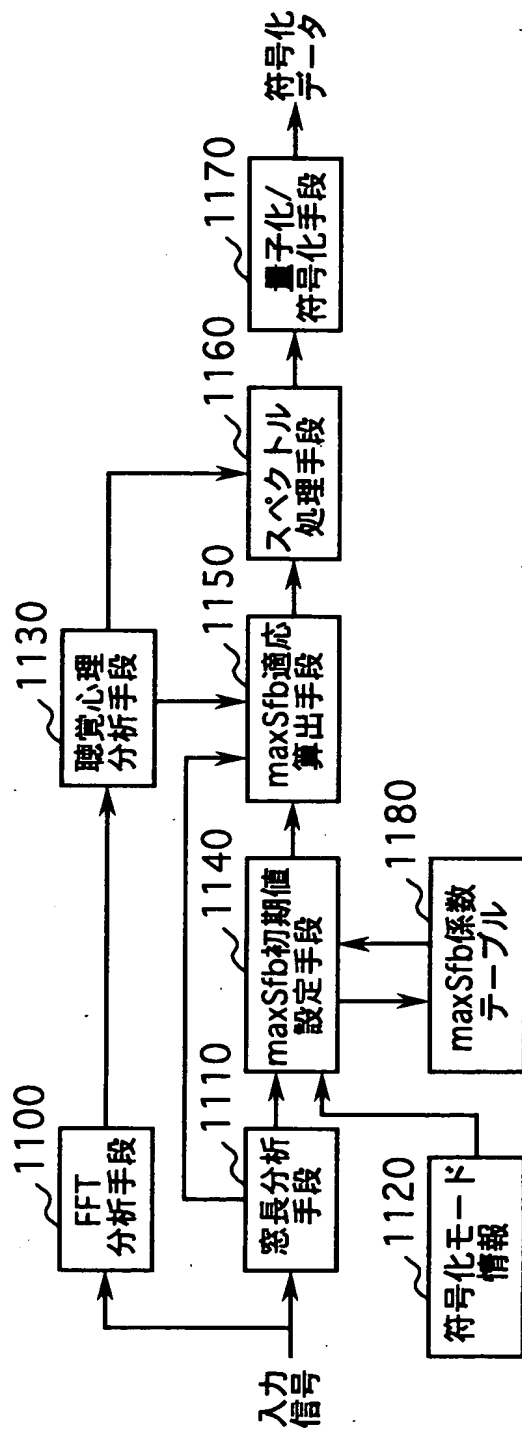
周波数 ヒット率	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	100000	100000	100000	50000	50000	50000	50000	50000	50000
96000	100000	100000	100000	100000	100000	50000	50000	50000	50000
80000	100000	100000	100000	100000	100000	100000	100000	100000	100000
64000	500000	500000	500000	500000	500000	500000	1000000	1000000	1000000

(b)

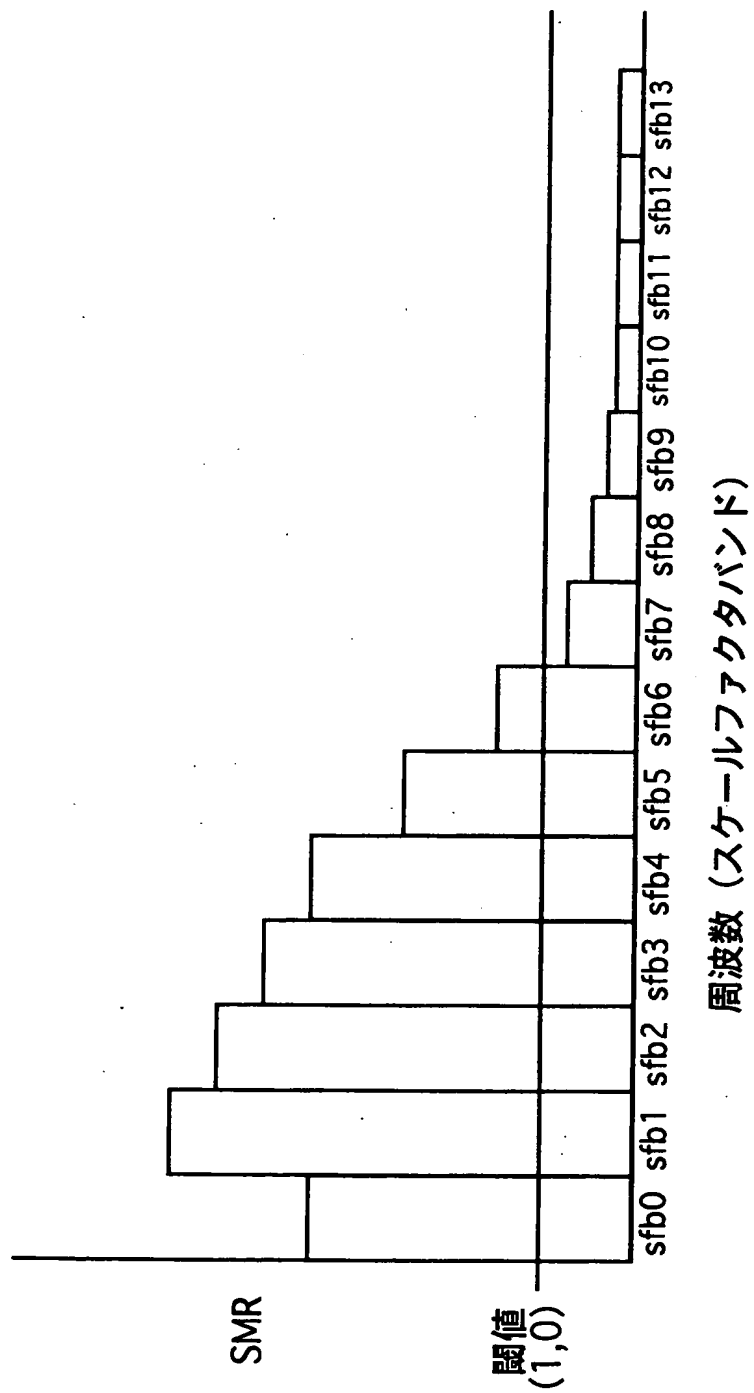
周波数 ヒット率	48000	44100	32000	24000	22050	16000	12000	11025	8000
128000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000
96000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000
80000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000
64000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000



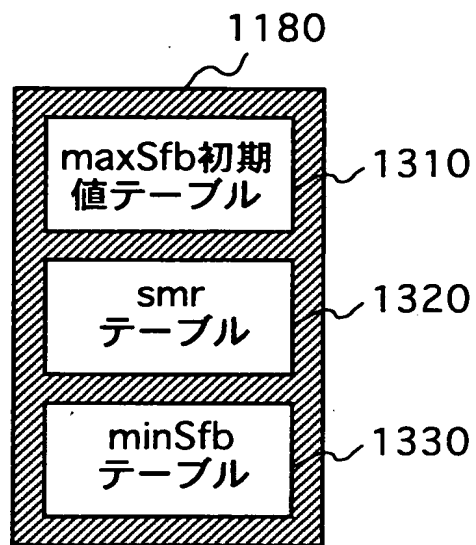
【図13】



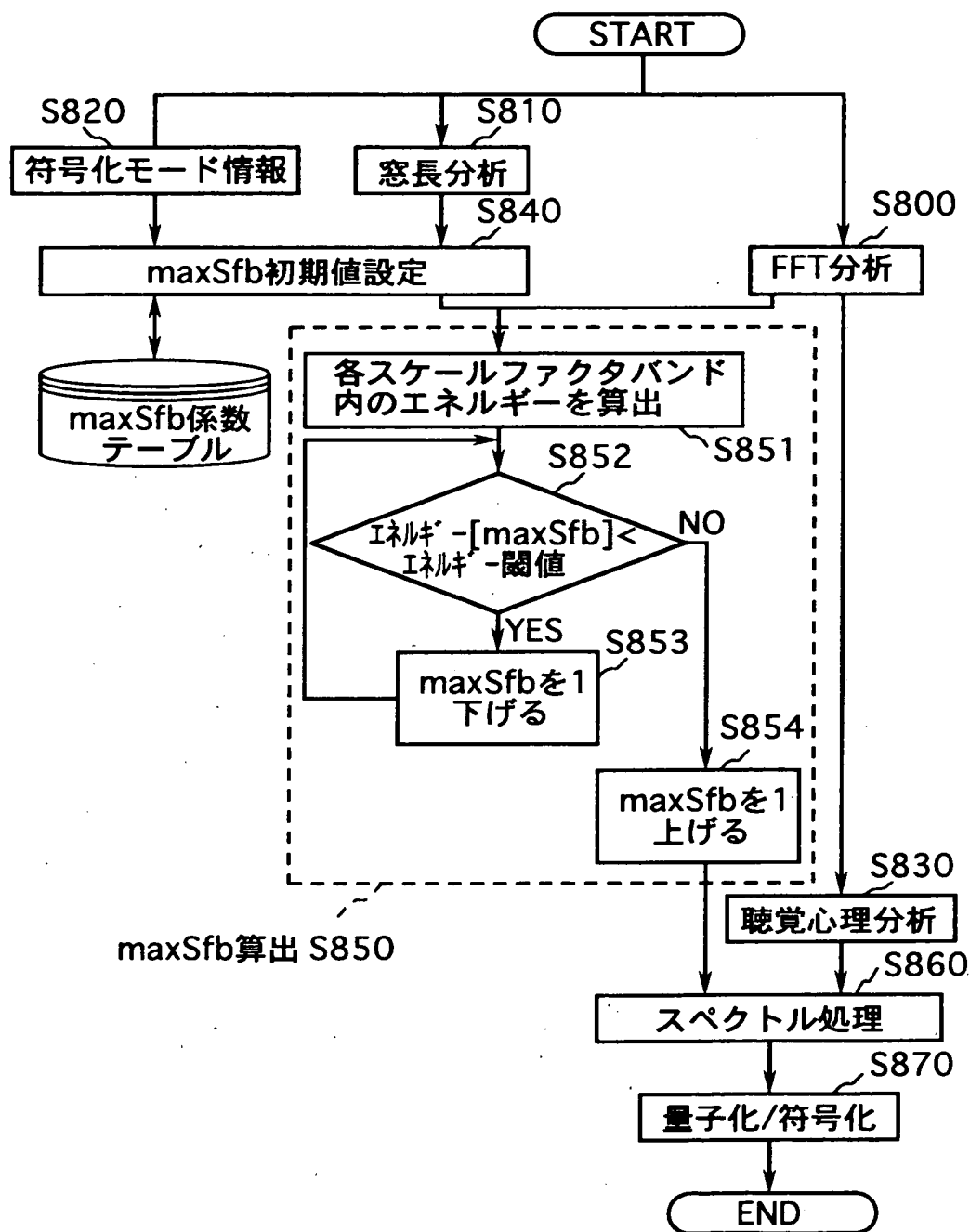
【図 1 4】



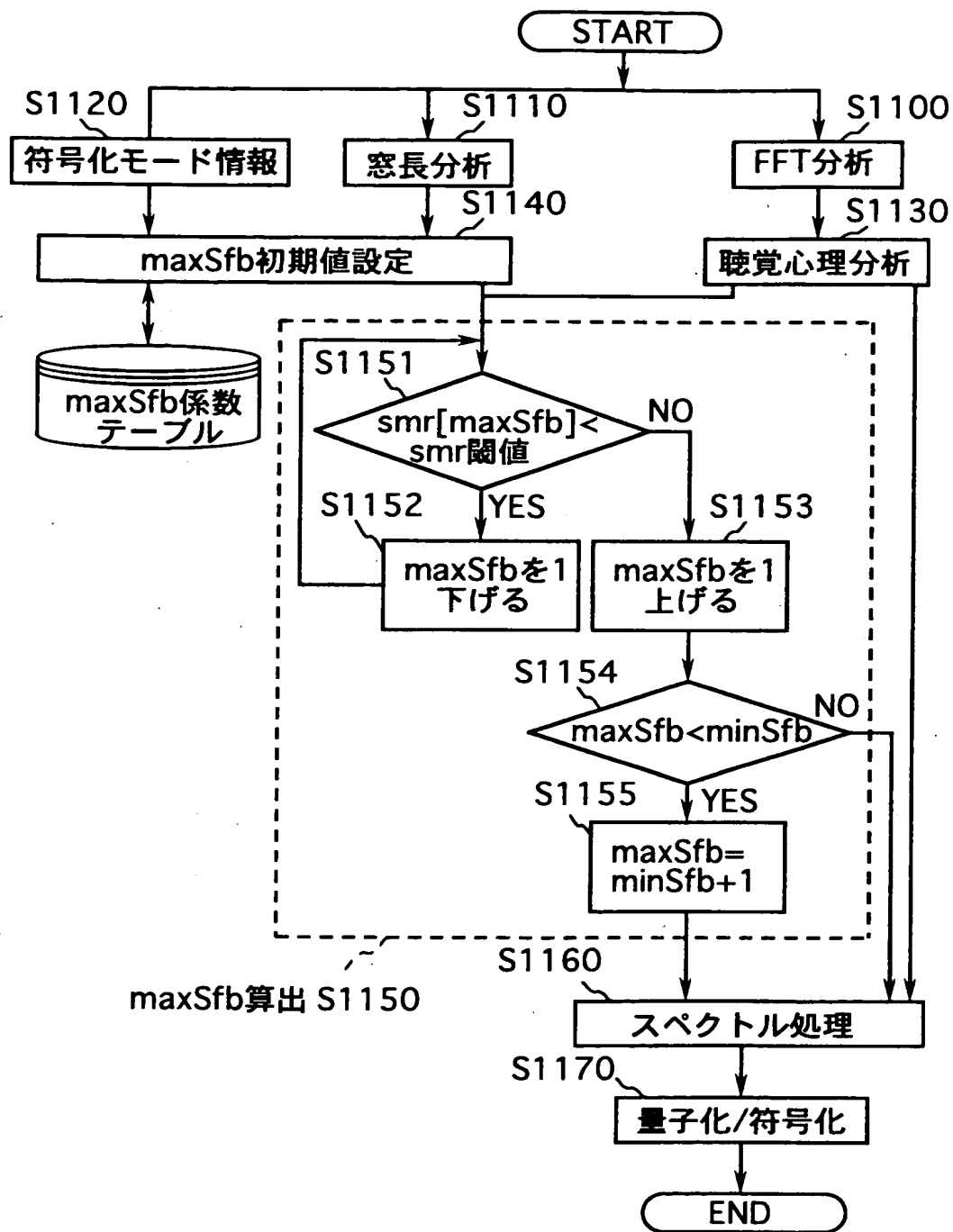
【図 1 5】



【図 16】



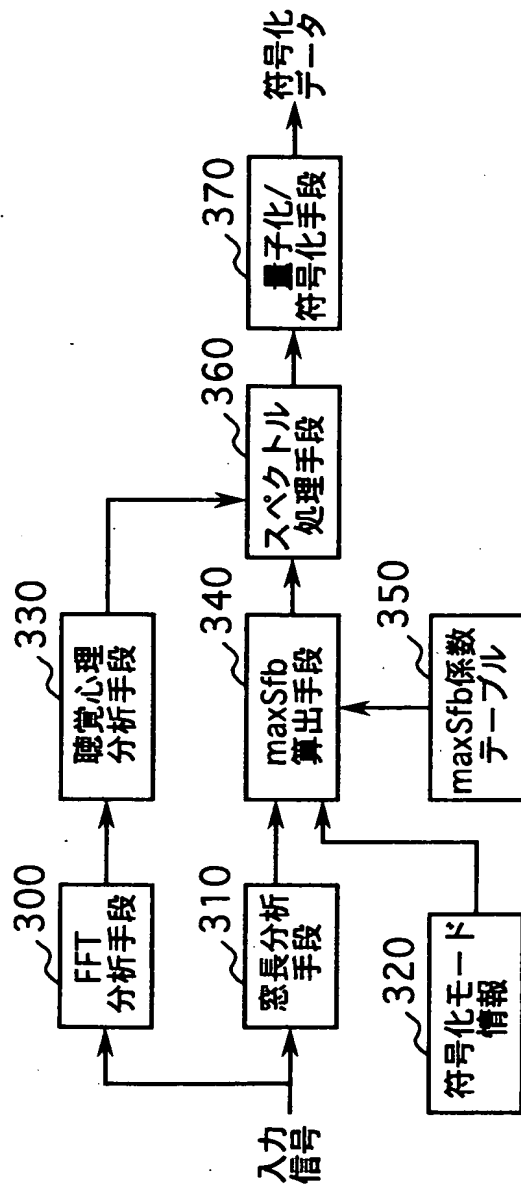
【図17】



【図18】

周波数	9600	88200	64000	48000	44100	32000	24000	22050	16000	12000	11025	8000
maxSfb (LONG)	49	49	47	48	49	51	47	47	43	43	43	40
maxSfb (SHORT)	12	12	12	13	14	13	15	15	15	15	15	15

【図 1 9】



【書類名】 要約書

【要約】

【課題】 M P E G 2 オーディオ符号化方式 A A C 方式等で符号化する際に、少ない演算量で、かつより高音質な符号化を実現することができる音声符号化方式を提供すること。

【解決手段】 窓長および符号化モードの指定情報に基づいて複数のスケールファクタバンドとその閾値情報を初期設定し、そのスケールファクタバンドのうち符号化する周波数帯域を特定する最大スケールファクタバンドを、入力信号の周波数分布を分析した結果および聴覚心理モデルに基づいて算出することにより、入力信号の周波数特性およびビットレートやサンプリング周波数等の符号化条件に応じて、符号化する周波数帯域を適応的に算出するように構成する。

【選択図】 図 1



出 願 人 履 歴 情 報

識別番号 [000005821]

1. 変更年月日	1990年 8月28日
[変更理由]	新規登録
住 所	大阪府門真市大字門真1006番地
氏 名	松下電器産業株式会社